Graphical models and their equations

Jan Draisma
Mathematical Institute
University of Bern
and Eindhoven University of Technology

Montpellier, 16 November 2017

Setting

$G = (V, E)$ finite, simple undirected graph

$\Omega_i, i \in V$ finite sets

P a probability distribution on the state space $\Omega := \prod_{i \in V} \Omega_i$

$X_i : \Omega \to \Omega_i$ the ith coordinate function

$A \subseteq V \rightsquigarrow$ probability vector X_A taking values in Ω_A.

Undirected graphical models

Setting

$G = (V, E)$ finite, simple undirected graph

$\Omega_i, i \in V$ finite sets

P a probability distribution on the state space $\Omega := \prod_{i \in V} \Omega_i$

$X_i : \Omega \to \Omega_i$ the ith coordinate function

$A \subseteq V \rightsquigarrow$ probability vector X_A taking values in Ω_A.

Conditional independence

$X_A \perp\!\!\!\perp X_B \mid X_C$ means: for each $x_C \in \Omega_C$ with $P(X_C = x_C) > 0$,

$P(X_A = x_A \wedge x_B = x_B \mid X_C = x_C) =$

$P(X_A = x_A \mid X_C = x_C) \cdot P(X_B = x_B \mid X_C = x_C)$.

Undirected graphical models

Setting

$G = (V, E)$ finite, simple undirected graph

$\Omega_i, i \in V$ finite sets

P a probability distribution on the state space $\Omega := \prod_{i \in V} \Omega_i$

$X_i : \Omega \to \Omega_i$ the ith coordinate function

$A \subseteq V \rightsquigarrow$ probability vector X_A taking values in Ω_A.

Conditional independence

$X_A \perp\!\!\!\perp X_B \mid X_C$ means: for each $x_C \in \Omega_C$ with $P(X_C = x_C) > 0$,

$P(X_A = x_A \wedge x_B = x_B \mid X_C = x_C) =$

$P(X_A = x_A \mid X_C = x_C) \cdot P(X_B = x_B \mid X_C = x_C)$.

Pairwise Markov properties from G

$X_i \perp\!\!\!\perp X_j \mid X_{V \setminus \{i,j\}}$ for $i \neq j$ with $\{i, j\} \notin E$.

Undirected graphical models

Example: Independence

$G =$ 

$$P((x_1, x_2, x_3)) = q_{x_1} r_{x_2} s_{x_3}$$

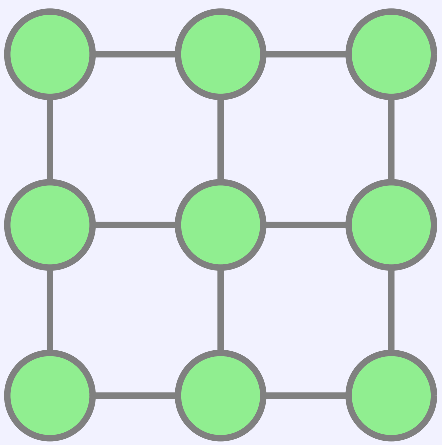
Example: Independence

$G =$

$$P((x_1, x_2, x_3)) = q_{x_1} r_{x_2} s_{x_3}$$

Example: Ising model

$G =$



$\Omega_i = \{-1, 1\}$ for all i

interaction parameters $c, d > 0$

$$P(x) = \frac{1}{Z} \cdot \left(\prod_{i \sim k, x_i = x_k} c\right) \cdot \left(\prod_{i \sim k, x_i \neq x_k} d\right)$$

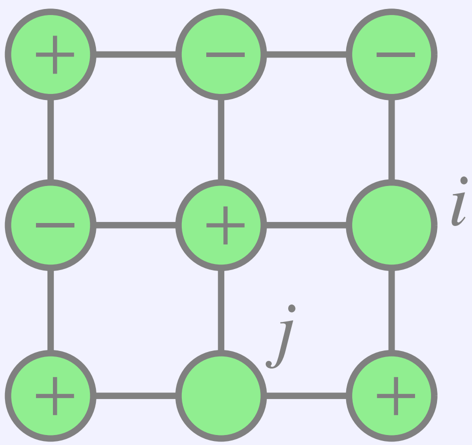
Undirected graphical models

Example: Independence

$G =$ 

$$P((x_1, x_2, x_3)) = q_{x_1} r_{x_2} s_{x_3}$$

Example: Ising model

$G =$ 

$\Omega_i = \{-1, 1\}$ for all i

interaction parameters $c, d > 0$

$$P(x) = \frac{1}{Z} \cdot \left(\prod_{i \sim k, x_i = x_k} c \right) \cdot \left(\prod_{i \sim k, x_i \neq x_k} d \right)$$

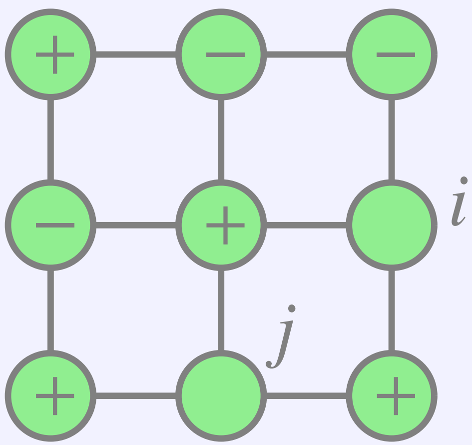
$$P((X_i, X_j) = (1, 1) \mid \cdots) = \frac{(c^2 d)(c^3)}{(c^2 d + c d^2)(c^3 + d^3)} = P(X_i = 1 \mid \cdots) \cdot P(X_j = 1 \mid \cdots)$$

Example: Independence

$G =$ 

$$P((x_1, x_2, x_3)) = q_{x_1} r_{x_2} s_{x_3}$$

Example: Ising model

$G =$ 

$\Omega_i = \{-1, 1\}$ for all i

interaction parameters $c, d > 0$

$$P(x) = \frac{1}{Z} \cdot \left(\prod_{i \sim k, x_i = x_k} c\right) \cdot \left(\prod_{i \sim k, x_i \neq x_k} d\right)$$

$$P((X_i, X_j) = (1,1) \mid \cdots) = \frac{(c^2 d)(c^3)}{(c^2 d + c d^2)(c^3 + d^3)} = P(X_i = 1 \mid \cdots) \cdot P(X_j = 1 \mid \cdots)$$

$\rightsquigarrow$ P satisfies all the pairwise Markov properties for G.

Hammersley-Clifford Theorem

Assume $P > 0$ on all of Ω. Then P satisfies all the pairwise Markov properties $\Leftrightarrow \exists$ *interaction parameters* $\theta_C \in \mathbb{R}_{>0}^{\Omega_C}$, where C runs through the maximal cliques of G, such that $P(x) = \prod_C \theta_C(x_C)$.

Hammersley-Clifford Theorem

Assume $P > 0$ on all of Ω. Then P satisfies all the pairwise Markov properties $\Leftrightarrow \exists$ *interaction parameters* $\theta_C \in \mathbb{R}_{>0}^{\Omega_C}$, where C runs through the maximal cliques of G, such that $P(x) = \prod_C \theta_C(x_C)$.

Example

• *Independence*: maximal cliques=vertices $\rightsquigarrow P(x) = \prod_{i \in V} \theta_i(x_i)$.

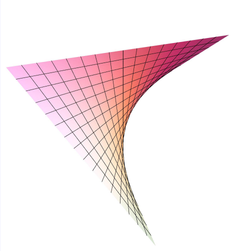
• *Ising*: maximal cliques are edges $\{i, k\}$, and $c = \theta_{ik}(-1,-1) = \theta_{ik}(1,1)$ and $d = \theta_{ik}(1,-1) = \theta_{ik}(-1,1)$ (up to normalisation).

Hammersley-Clifford Theorem

Assume $P > 0$ on all of Ω. Then P satisfies all the pairwise Markov properties $\Leftrightarrow \exists$ *interaction parameters* $\theta_C \in \mathbb{R}^{\Omega_C}_{>0}$, where C runs through the maximal cliques of G, such that $P(x) = \prod_C \theta_C(x_C)$.

Example

• *Independence*: maximal cliques=vertices $\rightsquigarrow P(x) = \prod_{i \in V} \theta_i(x_i)$.
• *Ising*: maximal cliques are edges $\{i,k\}$, and $c = \theta_{ik}(-1,-1) = \theta_{ik}(1,1)$ and $d = \theta_{ik}(1,-1) = \theta_{ik}(-1,1)$ (up to normalisation).



Working Definition

The *graphical model* M associated to (G,Ω) is the set of all positive probability distributions P on Ω as above—a semi-algebraic set in $\mathbb{R}^{\Omega}$ whose Zariski closure is a unirational variety in $\mathbb{R}^{\Omega}$. Denote by $\widehat{M}$ the cone over this closure.

Monomial parameterisation of $\widehat{M}$: $P(x) = \prod_C \theta_C(x_C)$.
Here we forget that the $P(x)$ must sum to 1 and must be positive. Hence the θ_C are unconstrained parameters. What *polynomial relations* among the $P(x)$ hold independently of the parameters θ_C?

Monomial parameterisation of $\widehat{M}$: $P(x) = \prod_C \theta_C(x_C)$.
Here we forget that the $P(x)$ must sum to 1 and must be positive. Hence the θ_C are unconstrained parameters. What *polynomial relations* among the $P(x)$ hold independently of the parameters θ_C?

Example: Independence
$P(x_1, x_2, x_3) = r_{x_1} s_{x_2} t_{x_3}$ satisfy the binomial equations $P(x_1, x_2, x_3)P(x_1', x_2', x_3) - P(x_1, x_2', x_3)P(x_1', x_2, x_3)$ and similar ones; these generate the ideal of all polynomial relations.

Monomial parameterisation of $\widehat{M}$: $P(x) = \prod_C \theta_C(x_C)$.
Here we forget that the $P(x)$ must sum to 1 and must be positive. Hence the θ_C are unconstrained parameters. What *polynomial relations* among the $P(x)$ hold independently of the parameters θ_C?

Example: Independence
$P(x_1, x_2, x_3) = r_{x_1} s_{x_2} t_{x_3}$ satisfy the binomial equations $P(x_1, x_2, x_3)P(x_1', x_2', x_3) - P(x_1, x_2', x_3)P(x_1', x_2, x_3)$ and similar ones; these generate the ideal of all polynomial relations.

Relevance for Fisher's exact test (Diaconis-Sturmfels, 1998)
These binomials give rise to a Markov chain that, starting from an Ω-contingency table $T \in \mathbb{Z}_{\geq 0}^{\Omega}$ samples such tables with the same *sufficient statistics*: counts of all patterns seen on cliques.

Example: Independence

$P(x_1, x_2, x_3) = r_{x_1} s_{x_2} t_{x_3}$ satisfy the binomial equations $P(x_1, x_2, x_3)P(x_1', x_2', x_3) - P(x_1, x_2', x_3)P(x_1', x_2, x_3)$ and similar ones; these generate the ideal of all polynomial relations—note that there are three orbits up to $\mathrm{Sym}(\Omega_1) \times \mathrm{Sym}(\Omega_2) \times \mathrm{Sym}(\Omega_3)$.

Example: Independence

$P(x_1,x_2,x_3) = r_{x_1}s_{x_2}t_{x_3}$ satisfy the binomial equations $P(x_1,x_2,x_3)P(x'_1,x'_2,x_3) - P(x_1,x'_2,x_3)P(x'_1,x_2,x_3)$ and similar ones; these generate the ideal of all polynomial relations—note that there are three orbits up to $\mathrm{Sym}(\Omega_1)\times\mathrm{Sym}(\Omega_2)\times\mathrm{Sym}(\Omega_3)$.

Much more generally:

Example: Independence

$P(x_1, x_2, x_3) = r_{x_1} s_{x_2} t_{x_3}$ satisfy the binomial equations $P(x_1, x_2, x_3)P(x_1', x_2', x_3) - P(x_1, x_2', x_3)P(x_1', x_2, x_3)$ and similar ones; these generate the ideal of all polynomial relations—note that there are three orbits up to $\mathrm{Sym}(\Omega_1) \times \mathrm{Sym}(\Omega_2) \times \mathrm{Sym}(\Omega_3)$.

Much more generally:

Independent Set Theorem (Hillar-Sullivant, 2012)

If $A \subseteq V$ is an independent set in G, then the ideal of $\widehat{M}$ is generated by boundedly many $\prod_{i \in A} \mathrm{Sym}(\Omega_i)$-orbits of binomials as $|\Omega_i| \to \infty$ for all $i \in A$.

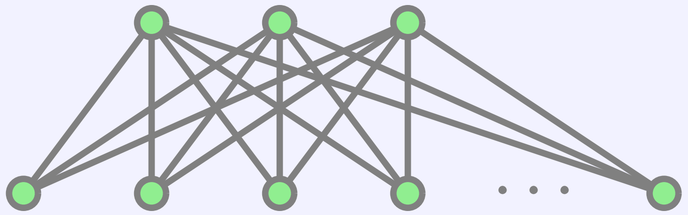
Example: Independence
$P(x_1, x_2, x_3) = r_{x_1} s_{x_2} t_{x_3}$ satisfy the binomial equations $P(x_1, x_2, x_3)P(x'_1, x'_2, x_3) - P(x_1, x'_2, x_3)P(x'_1, x_2, x_3)$ and similar ones; these generate the ideal of all polynomial relations—note that there are three orbits up to $\mathrm{Sym}(\Omega_1) \times \mathrm{Sym}(\Omega_2) \times \mathrm{Sym}(\Omega_3)$.

Much more generally:

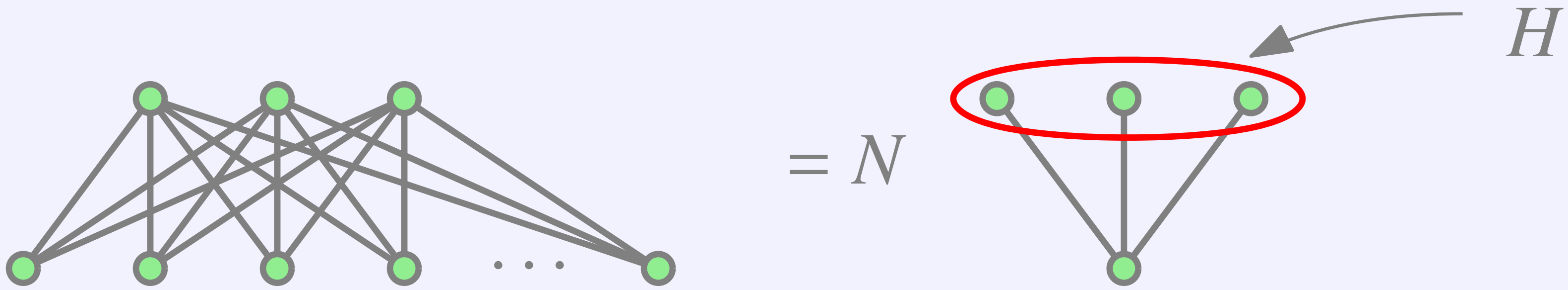
Independent Set Theorem (Hillar-Sullivant, 2012)
If $A \subseteq V$ is an independent set in G, then the ideal of $\widehat{M}$ is generated by boundedly many $\prod_{i \in A} \mathrm{Sym}(\Omega_i)$-orbits of binomials as $|\Omega_i| \to \infty$ for all $i \in A$.

Crucial fact: maps $f_i : \Omega_i \to \Omega'_i, i \in A$ together yield a linear map $\mathbb{R}^{\Omega'} \to \mathbb{R}^{\Omega}$ sending $\widehat{M}'$ into $\widehat{M}$; use this to pull back equations.

Example (Rauh-Sullivant, 2014)
If G is $K_{3,N}$, and all state spaces are $\{0, 1\}$, then the ideal of $\widehat{M}$ is generated by binomials of degree ≤ 12, independently of N.

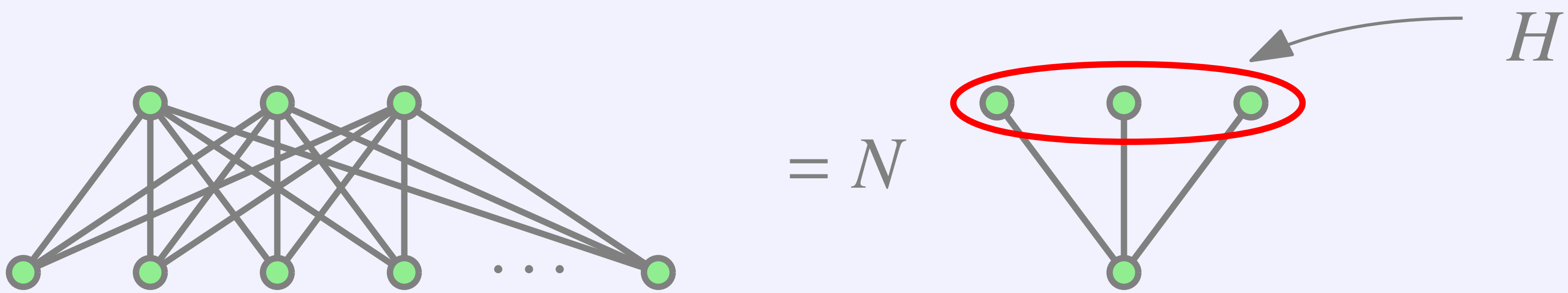


Example (Rauh-Sullivant, 2014)
If G is $K_{3,N}$, and all state spaces are $\{0,1\}$, then the ideal of $\widehat{M}$ is generated by binomials of degree ≤ 12, independently of N.



Construction: $G_1, \ldots, G_k$ finite graphs with a common induced subgraph $H \rightsquigarrow s_1 G_1 +_H \cdots +_H s_k G_k$ obtained from disjoint copies of the G_j by identifying their instances of H.

Example (Rauh-Sullivant, 2014)
If G is $K_{3,N}$, and all state spaces are $\{0, 1\}$, then the ideal of $\widehat{M}$ is generated by binomials of degree ≤ 12, independently of N.



Construction: $G_1, \ldots, G_k$ finite graphs with a common induced subgraph $H \rightsquigarrow s_1 G_1 +_H \cdots +_H s_k G_k$ obtained from disjoint copies of the G_j by identifying their instances of H.

Theorem (D-Oosterhof, 2016)
Fixing state spaces for the vertices of each G_j, compatible with H, the ideal of $\widehat{M}(s_1 G_1 +_H \cdots +_H s_k G_k)$ is generated in bounded degree uniformly in the s_j.

Growing graphs

Crucial fact: Suppose that G has vertex set $A \sqcup B$, where A is the vertex set of H; so G has state space $\Omega_A \times \Omega_B$.

Growing graphs

Crucial fact: Suppose that G has vertex set $A \sqcup B$, where A is the vertex set of H; so G has state space $\Omega_A \times \Omega_B$.

Then sG has the vertex set $A \sqcup ([s] \times B)$ and state space $\Omega(s) := \Omega_A \times \Omega_B^s$. Any map $f : [s] \to [r]$ yields a map $\Omega(r) \to \Omega(s)$ and a linear map $\mathbb{R}^{\Omega(s)} \to \mathbb{R}^{\Omega(r)}$, which turns out to map $\widehat{M}(s)$ in the former space into $\widehat{M}(r)$ in the latter space.

Crucial fact: Suppose that G has vertex set $A \sqcup B$, where A is the vertex set of H; so G has state space $\Omega_A \times \Omega_B$.

Then sG has the vertex set $A \sqcup ([s] \times B)$ and state space $\Omega(s) := \Omega_A \times \Omega_B^s$. Any map $f : [s] \to [r]$ yields a map $\Omega(r) \to \Omega(s)$ and a linear map $\mathbb{R}^{\Omega(s)} \to \mathbb{R}^{\Omega(r)}$, which turns out to map $\widehat{M}(s)$ in the former space into $\widehat{M}(r)$ in the latter space.

Thus $\widehat{M}$ is a *variety over the category* **Fin** *of finite sets*. We show that its ambient space is a Noetherian **Fin**-variety.

(The Independent Set Theorem concerns a $\mathbf{Fin}^{\mathrm{op}}$-variety!)

Open: What happens if both state spaces and graphs grow?

Mixtures

Definition

Given two models M_1 and M_2 in R^Ω, their *mixture* is the set $\{\lambda P + (1-\lambda)Q \mid P \in M_1, Q \in M_2, \lambda \in [0,1]\}$.

Definition
Given two models M_1 and M_2 in R^Ω, their *mixture* is the set $\{\lambda P + (1-\lambda)Q \mid P \in M_1, Q \in M_2, \lambda \in [0,1]\}$.

Example (Independence)
The mixture of two copies of independence is the set of all $|\Omega_1| \times |\Omega_2| \times |\Omega_3|$-tensors of *nonnegative rank* at most two whose entries add up to one.

Definition
Given two models M_1 and M_2 in R^Ω, their *mixture* is the set $\{\lambda P + (1-\lambda)Q \mid P \in M_1, Q \in M_2, \lambda \in [0,1]\}$.

Example (Independence)
The mixture of two copies of independence is the set of all $|\Omega_1| \times |\Omega_2| \times |\Omega_3|$-tensors of *nonnegative rank* at most two whose entries add up to one.

Theorem (Allman-Rhodes-Sturmfels-Zwiernik, 2013)
An $m_1 \times m_2 \times m_3$-tensor P with entries in $\mathbb{R}_{\geq 0}$ has nonnegative rank at most two if and only if P has rank at most two and is moreover *(log-)supermodular*: $P(x_1, x_2, x_3)P(y_1, y_2, y_3) \leq P(u_1, u_2, u_3)P(z_1, z_2, z_3)$ if $\{x_r, y_r\} = \{u_r, z_r\}$ and $u_r \leq z_r$ for all r, or in the $\mathrm{Sym}(\Omega_1) \times \mathrm{Sym}(\Omega_2) \times \mathrm{Sym}(\Omega_3)$-orbit of such a tensor.

Boundedness of equations of some mixtures

Forget again about inequalities and summing up to 1.

Easy fact:
If G is a disjoint union of cliques, then $\widehat{M}$ is not only stable under $\prod_{i \in V} \mathrm{Sym}(\Omega_i)$, but even under $\prod_{i \in V} \mathrm{GL}_{\Omega_i}$. Hence the same holds for mixtures $\widehat{M}_1 + \widehat{M}_2 = \overline{\{P + Q \mid P \in \widehat{M}_1, Q \in \widehat{M}_2\}}$ coming from graphs G_1, G_2 that are unions of cliques.

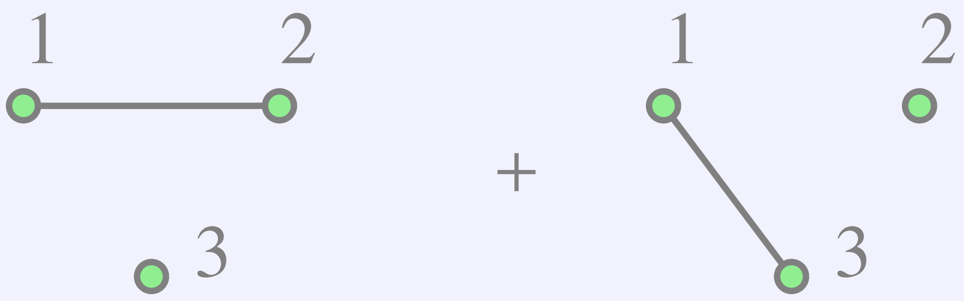
Forget again about inequalities and summing up to 1.

Easy fact:
If G is a disjoint union of cliques, then $\widehat{M}$ is not only stable under $\prod_{i \in V} \mathrm{Sym}(\Omega_i)$, but even under $\prod_{i \in V} \mathrm{GL}_{\Omega_i}$. Hence the same holds for mixtures $\widehat{M}_1 + \widehat{M}_2 = \overline{\{P + Q \mid P \in \widehat{M}_1, Q \in \widehat{M}_2\}}$ coming from graphs G_1, G_2 that are unions of cliques.

Theorem (D, 2017)
For any fixed k, a closed subvariety in a tensor product $W_1 \otimes \cdots \otimes W_k$ of vector spaces that depends functorially on $W_1, \ldots, W_k$ is defined by finitely many equations up to $\prod_i \mathrm{GL}(W_i)$, independently of the dimensions of the W_i.

A mixture challenge

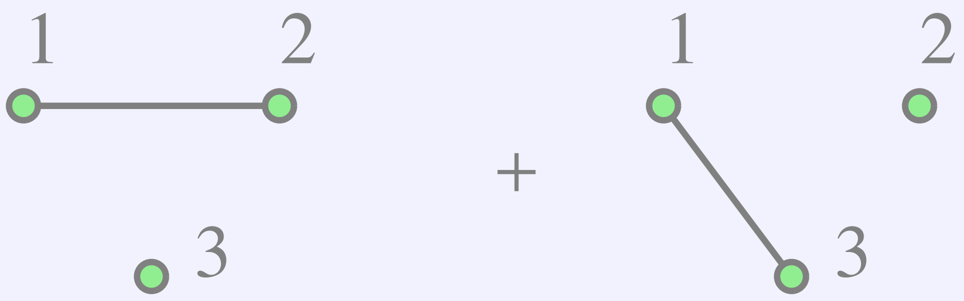
A mixture challenge



$$M = \{P(x_1, x_2, x_3) = \lambda \frac{1}{Z(c,d)} c_{x_1,x_2} d_{x_3} + (1 - \lambda) \frac{1}{Z'(e,f)} e_{x_1,x_3} f_{x_2}\}$$

Here: $\lambda \in [0, 1]$, $c \in \mathbb{R}_{>0}^{\Omega_1 \times \Omega_2}$, $d \in \mathbb{R}_{>0}^{\Omega_3}$, $e \in \mathbb{R}_{>0}^{\Omega_1 \times \Omega_3}$, $f \in \mathbb{R}_{>0}^{\Omega_2}$

A mixture challenge



$$M = \{P(x_1, x_2, x_3) = \lambda \frac{1}{Z(c,d)} c_{x_1,x_2} d_{x_3} + (1-\lambda) \frac{1}{Z'(e,f)} e_{x_1,x_3} f_{x_2}\}$$

Here: $\lambda \in [0,1]$, $c \in \mathbb{R}_{>0}^{\Omega_1 \times \Omega_2}$, $d \in \mathbb{R}_{>0}^{\Omega_3}$, $e \in \mathbb{R}_{>0}^{\Omega_1 \times \Omega_3}$, $f \in \mathbb{R}_{>0}^{\Omega_2}$

Challenge: Find a quantifier-free description of M!

Oosterhof found polynomial equations cutting out $\widehat{M}$ of degrees 3 and 6: certain 2×2-determinants of 3×3-determinants.

There's a beautiful relation with matrix spaces of rank two!

Many other topics

• equations for phylogenetic tree models (Casanellas et al, Allman-Rhodes, Sturmfels-Sullivant, Michałek et al, ...)
• determinantal equations for Gaussian graphical models (Sullivant-Talaska-D)
• identifiability of Gaussian graphical models (Foygel-Drton-D)
• (non-)singularity of hypersurfaces defined by conditional independence statements for Gaussian graphical models (Lin-Uhler-Sturmfels-Bühlmann)
• twisted commutative algebras (Sam-Snowden)

$\vdots$

Many other topics

• equations for phylogenetic tree models (Casanellas et al, Allman-Rhodes, Sturmfels-Sullivant, Michałek et al, . . .)
• determinantal equations for Gaussian graphical models (Sullivant-Talaska-D)
• identifiability of Gaussian graphical models (Foygel-Drton-D)
• (non-)singularity of hypersurfaces defined by conditional independence statements for Gaussian graphical models (Lin-Uhler-Sturmfels-Bühlmann)
• twisted commutative algebras (Sam-Snowden)

$\vdots$

Thank you!